

21

Evaluación conductual directa: qué registrar y cómo hacerlo

SUPONGAMOS que ha elegido una conducta concreta para modificarla. ¿Cómo se mide, evalúa, o calcula directamente esa conducta? Como mencionamos en el Capítulo 20, siempre que sea posible, los modificadores de conducta suelen preferir las medidas directas frente a las indirectas. Al medir la conducta directamente, hay seis características a considerar: topografía, cantidad, intensidad, control de los estímulos, latencia y cualidad.

CARACTERÍSTICAS DE LA CONDUCTA A REGISTRAR

Topografía

Como indicamos en el Capítulo 10, la *topografía* hace referencia a la forma de la respuesta concreta, es decir, consiste en una descripción de los movimientos específicos implicados. Imaginemos que una profesora quiere moldear en una alumna con trastorno del desarrollo el gesto de levantar un brazo para obtener atención en clase. La profesora podría identificar maneras de levantar el brazo, de calidad creciente, tal como se muestra en la Tabla 21-1, y proceder al moldeado con un programa desde el paso 1 al 6.

Los dibujos suelen ser útiles como clave para ayudar a los observadores a identificar variaciones en la topografía de una respuesta. Uno de los autores desarrolló una lista de comprobación detallada con dibujos instigadores para evaluar el estilo de las brazadas de nadadores jóvenes. La Figura 21-1 muestra esta lista de comprobación para la brazada de espalda.

Cantidad

Dos medidas comunes de la cantidad de una conducta son la frecuencia y la duración.

Frecuencia: la *frecuencia* se refiere al número de episodios de una conducta que tienen lugar en un periodo de tiempo concreto. Los términos *tasa* o *frecuencia*

Tabla 21-1. Distintas maneras de levantar el brazo, de calidad pobre a buena

1. Mientras está sentada con los dos brazos apoyados en la mesa,	la alumna levanta el brazo de forma que la mano y el antebrazo quedan a unos centímetros de la mesa
2. Mientras está sentada con los dos brazos apoyados en la mesa,	la alumna levanta un brazo de manera que queda aproximadamente a la altura de su barbilla
3. Mientras está sentada con los dos brazos apoyados en la mesa,	la alumna levanta un brazo de forma que queda aproximadamente a la altura de sus ojos
4. Mientras está sentada con los dos brazos apoyados en la mesa,	la alumna levanta un brazo de manera que la mano queda un poco por encima de su cabeza
5. Mientras está sentada con los dos brazos apoyados en la mesa,	la alumna levanta un brazo de forma que la mano queda apuntando hacia arriba unos centímetros por encima de la cabeza, pero el codo sigue doblado
6. Mientras está sentada con los dos brazos apoyados en la mesa,	la alumna levanta un brazo de manera tal que queda recto apuntando hacia arriba

se suelen usar indistintamente. Si quisiera, por ejemplo, mejorar el rendimiento en los entrenamientos de jóvenes patinadores, podría examinar la frecuencia con que realizan saltos y piruetas durante los entrenamientos. Esta fue la aproximación adoptada por Michelle Hume, una entrenadora de patinaje artístico del club de patinaje St. Anne's Figure Skating en Manitoba (Hume y cols., 1985; véase **Nota 1**). La entrenadora Hume definió los saltos y piruetas de forma que los estudiantes que observaban los entrenamientos pudieran decidir cuáles de esas respuestas tenían lugar. Un *salto* estaba definido como cualquier ocasión en que los patinadores saltaban en el aire de forma que los dos patines se levantaban del hielo, hacían por lo menos un giro completo en el aire, y aterrizaban sobre un pie, de espaldas y sin caerse. Una *vuelta* estaba definida como el giro sobre un patín durante un mínimo de tres revoluciones mientras se mantiene el equilibrio y la postura. Cuando los observadores supieron qué conductas buscar, el siguiente paso de Hume fue tomar una línea de base de cuántos saltos y giros realizaban los patinadores durante los entrenamientos. Los observadores anotaron los datos que se muestran en la Figura 21-2.

En muchas situaciones, una persona no tiene ayudantes o carece de tiempo para coger papel y lápiz y ponerse a tabular. Por suerte, existen otras formas de evaluación cuantitativa que requieren muy poco tiempo. Un método sería usar un contador, como el tipo de reloj de muñeca relativamente barato utilizado por los jugadores de golf para registrar su puntuación. Con estos contadores se registran hasta 99 episodios apretando simplemente un botón por cada ocasión en que aparezca la conducta. Otra técnica sencilla de registro es transferir ítems, como bolitas, de un bolsillo a otro. Al final de la sesión, o al final del día, dependiendo de la actividad concreta que esté registrando, se hace el arqueo y se anota la cantidad de las bolitas en el segundo bolsillo. Se puede usar también una calculadora electrónica y apretar la tecla «+1» cada vez que ocurra la acción para que la calculadora mantenga el registro del total. Los ordenadores portátiles permiten recoger más de una actividad, o la actividad de varias personas, a la vez que almacenan los tiempos en que se han producido (Paggeot, Kvale, Mace y Sharkey, 1988; Repp, Karsh, Felce y Ludewig, 1989). Casi siempre

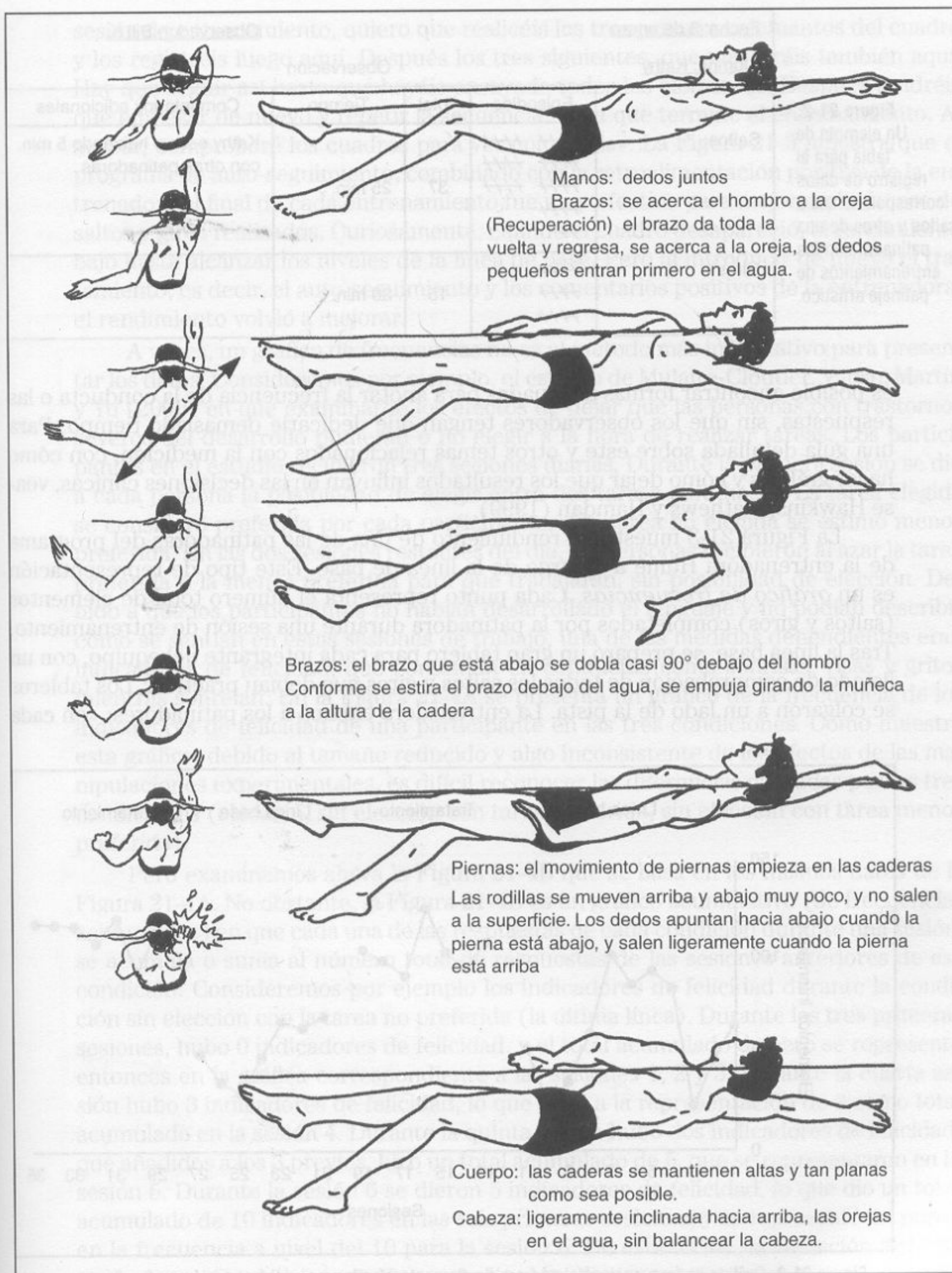


Figura 21-1. Lista de comprobación para los movimientos de natación de estilo espalda.

Figura 21-2.
Un ejemplo de
tabla para el
registro de datos
correspondientes a
saltos y giros de una
patinadora en los
entrenamientos de
patinaje artístico.

Fecha: 3 de enero		Observador: Bill K.		
Alumna: Kathy		Observación		
	Episodios	Total	Tiempo	Comentarios adicionales
Saltos	//// //// //// //// ##	37	25 min.	Kathy estuvo hablando 5 min. con otras patinadoras
Giros	//// //// ////	15	20 min.	

es posible encontrar formas adecuadas para anotar la frecuencia de la conducta o las respuestas, sin que los observadores tengan que dedicarle demasiado tiempo. Para una guía detallada sobre este y otros temas relacionados con la medición, con cómo hacer gráficas y cómo dejar que los resultados influyan en las decisiones clínicas, véase Hawkins, Mathews y Hamdan (1999).

La Figura 21-3 muestra el rendimiento de una de las patinadoras del programa de la entrenadora Hume a lo largo de la línea de base. Este tipo de representación es un *gráfico de frecuencias*. Cada punto representa el número total de elementos (saltos y giros) completados por la patinadora durante una sesión de entrenamiento. Tras la línea base, se preparó un gran tablero para cada integrante del equipo, con un listado de comprobación de todos los saltos y giros que debían practicar. Los tableros se colgaron a un lado de la pista. La entrenadora explicó a los patinadores: «en cada

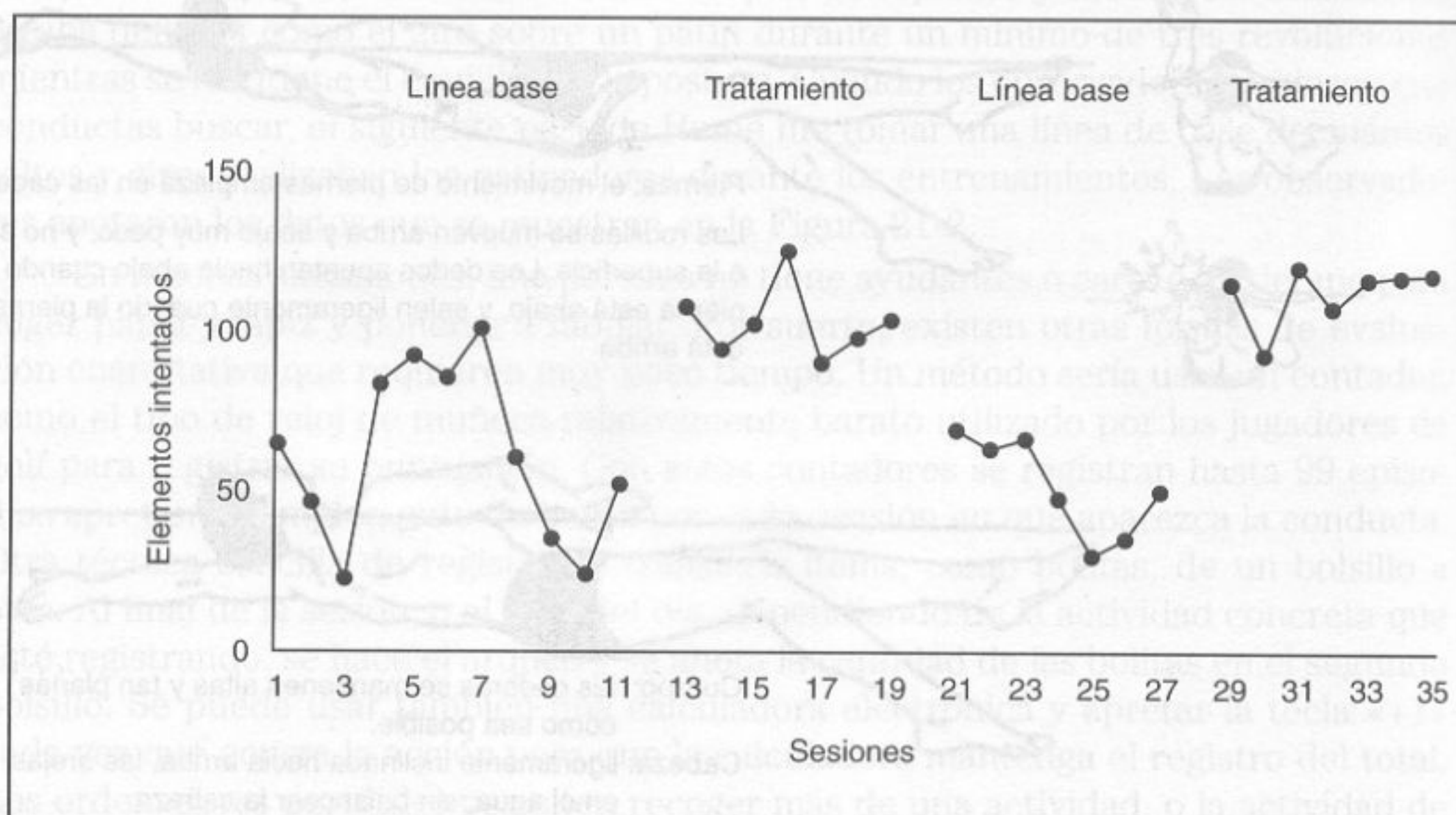


Figura 21-3. Gráfico de frecuencias del número de elementos (saltos y giros) por sesión realizados por una patinadora durante la línea base y el tratamiento (auto-registro y retroalimentación de la entrenadora).

sesión de entrenamiento, quiero que realicéis los tres primeros elementos del cuadro y los registréis luego aquí. Después los tres siguientes, que marcaréis también aquí. Hay que seguir así hasta que hayáis ensayado todos los elementos. Después tendréis que empezar de nuevo y repetir la secuencia hasta que termine el entrenamiento. Al acabar, comprobaré los cuadros para ver cómo vais». La Figura 21-3 muestra que el programa de auto-seguimiento, combinado con la retroalimentación positiva de la entrenadora al final de cada entrenamiento fue muy efectivo para mejorar el número de saltos y giros realizados. Curiosamente, cuando el cuadro desapareció, el rendimiento bajó hasta alcanzar los niveles de la línea de base. Pero al introducir de nuevo el tratamiento, es decir, el auto-seguimiento y los comentarios positivos de la entrenadora, el rendimiento volvió a mejorar.

A veces, un gráfico de frecuencias no es el método más informativo para presentar los datos. Consideremos por ejemplo, el estudio de Mulaire-Cloutier, Vause, Martin y Yu (2000) en que examinaron los efectos de dejar que las personas con trastornos severos del desarrollo pudieran o no elegir a la hora de realizar tareas. Los participantes en el estudio recibieron tres sesiones diarias. Durante la primera sesión se dio a cada persona la posibilidad de elegir entre dos tareas para hacer. La tarea elegida se consideró preferida por cada participante, y la tarea no elegida se estimó menos preferida. En las dos sesiones restantes del día, las personas recibieron al azar la tarea preferida o la menos preferida para que trabajaran, sin posibilidad de elección. Debido a que los participantes no habían desarrollado el lenguaje y no podían describir cómo se sentían en estas sesiones de trabajo, una de las medidas dependientes eran indicadores de felicidad, que consistían en conductas como sonrisas, risas y gritos mientras sonreían. En la Figura 21-4A se presenta un gráfico de la frecuencia de los indicadores de felicidad de una participante en las tres condiciones. Como muestra esta gráfica, debido al tamaño reducido y algo inconsistente de los efectos de las manipulaciones experimentales, es difícil reconocer las diferencias causadas por las tres condiciones (elección, sin elección con tarea preferida, sin elección con tarea menos preferida).

Pero examinemos ahora la Figura 21-4B que se basa en los mismos datos de la Figura 21-4A. No obstante, la Figura 21-4B es un gráfico acumulativo (de frecuencias acumuladas) en que cada una de las respuestas de cada condición durante una sesión, se acumula o suma al número total de respuestas de las sesiones anteriores de esa condición. Consideremos por ejemplo los indicadores de felicidad durante la condición sin elección con la tarea no preferida (la última línea). Durante las tres primeras sesiones, hubo 0 indicadores de felicidad, y el total acumulado de cero se representó entonces en la gráfica correspondiente a las sesiones 1, 2 y 3. Durante la cuarta sesión hubo 3 indicadores de felicidad, lo que llevó a la representación de 3 como total acumulado en la sesión 4. Durante la quinta sesión hubo dos indicadores de felicidad, que añadidos a los 3 previos, hizo un total acumulado de 5, que se representaron en la sesión 5. Durante la sesión 6 se dieron 5 indicadores de felicidad, lo que dio un total acumulado de 10 indicadores en las seis primeras sesiones, y se representó un punto en la frecuencia a nivel del 10 para la sesión 6. De esta forma, la actuación durante cualquier sesión de la condición se añadía al rendimiento total de las sesiones previas de esa condición y luego se representaba en un registro acumulado.

Se puede advertir en la gráfica acumulada la inclinación de la pendiente de la línea que indica cuántas respuestas ocurrieron durante un periodo concreto de tiempo. En otras palabras, la inclinación de la línea proporciona una idea de la tasa de respuestas. La baja pendiente de la línea durante la condición en que, sin poder elegir, le había correspondido por azar la tarea no preferida indica una tasa muy baja de gestos de felicidad en la participante. La pendiente más elevada durante la condición de elección muestra la tasa más alta de indicadores de felicidad. Hay que destacar otra característica de las gráficas de frecuencias acumuladas: la línea nunca puede bajar. Si una participante no está actuando, como en las primeras tres sesiones de la condición

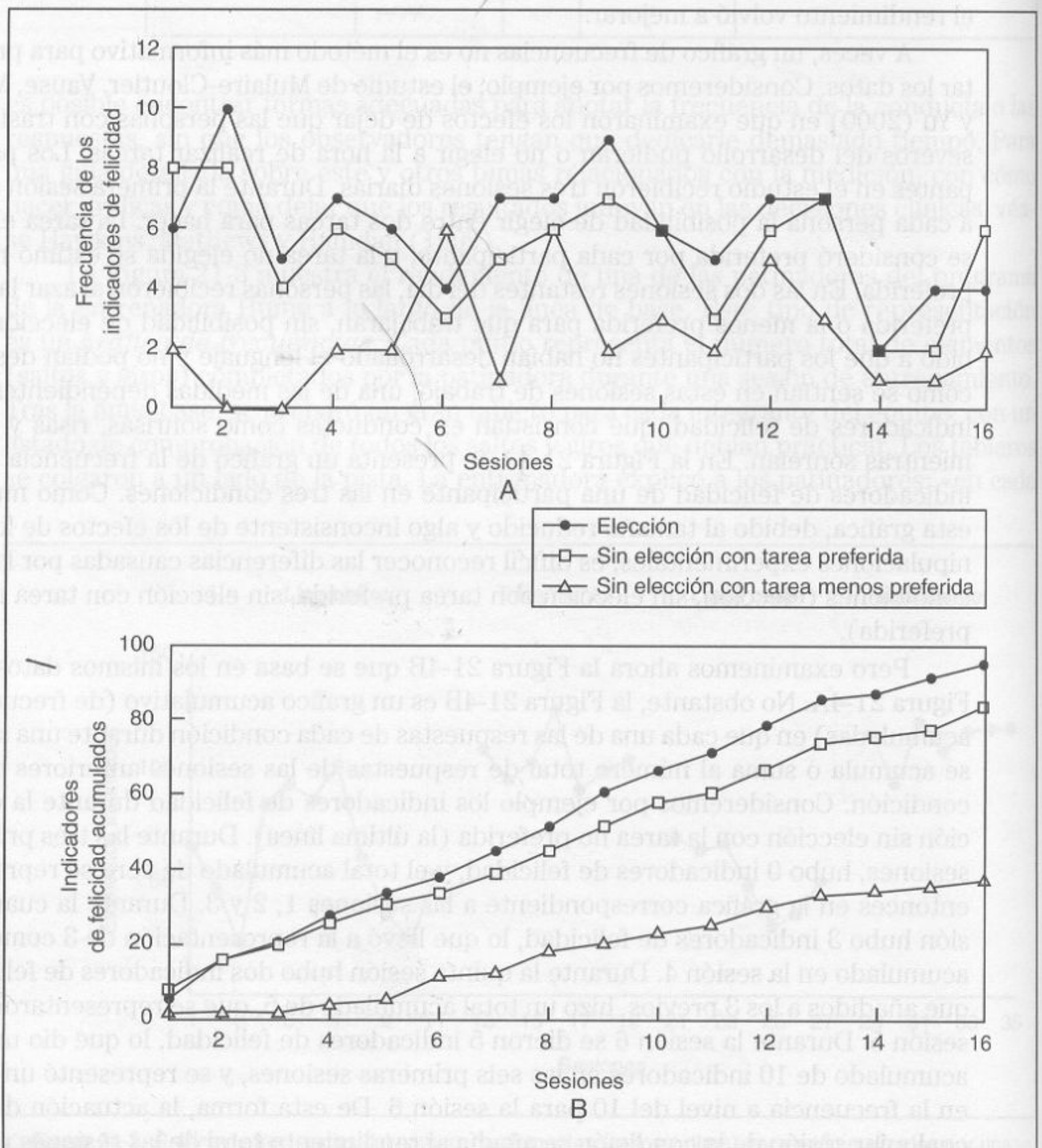


Figura 21-4. Un gráfico de frecuencias (A) y un gráfico de frecuencias acumuladas (B) de los mismos datos.

sin elección con tarea no preferida, no hay respuestas que se puedan acumular con lo que había previamente, y por eso la línea es plana.

Se suele preferir una gráfica acumulada frente a una gráfica de frecuencias directas cuando se comparan dos o más conductas o condiciones concurrentes, y cuando las diferencias entre sesiones son muy pequeñas. Es evidente que las diferencias en frecuencia de los indicadores de felicidad entre las tres condiciones (elección, no elección con tarea preferida y no elección con tarea menos preferida) son difíciles de detectar cuando se representan en gráficos de frecuencias directas en la Figura 21-4A. Sin embargo, cuando estos mismos datos se representan acumulados en la Figura 21-4B, la dispersión gradual de los resultados acumulados muestra un claro efecto de la condición de elección y en la tarea preferida sobre la no preferida.

A veces es posible diseñar una hoja que registre los datos en crudo y sirva a la vez como gráfica final. Consideremos el caso ficticio de Jackie, una chica que solía decir tacos a la profesora y a sus ayudantes. Para tratar este problema, supongamos que se diseñó el siguiente programa: cada vez que la profesora o sus ayudantes observaban un insulto, durante la línea base, ignoraban a Jackie, se dirigían a una tabla en el escritorio y ponían una X en el lugar apropiado. El gráfico resultante se muestra en la Figura 21-5.

Se observa en la Figura 21-5, que los exabruptos se registraban a lo largo del eje vertical, y los días en el eje horizontal. Cada vez que la niña decía un taco, el personal simplemente añadía una X en el día correspondiente al número de Xs que ya había acumuladas. El gráfico muestra claramente que ignorar la conducta de Jackie no tuvo ningún efecto. Esto pudo ocurrir si la conducta era reforzada por los demás estudiantes, aunque no lo fuera por la profesora y sus ayudantes. El gráfico también muestra que cuando Jackie empezó un programa de tratamiento en que recibía reforzamiento al final de cada periodo de quince minutos en que no dijera tacos (programa de refuerzo diferencial de respuesta cero, RD0, quince minutos), las palabras malsonantes se redujeron hasta desaparecer.

Los episodios de conducta que se registran en términos de frecuencia, como saltar o hacer giros en el caso de los patinadores artísticos, son ejemplos de una actividad distinta, separada, individual, de la que es fácil llevar el cómputo durante un periodo de tiempo concreto. Los modificadores de conducta han registrado la frecuencia de actividades como decir una palabra concreta, gritar insultos, lanzar objetos, solucionar problemas aritméticos, masticar bocados de comida, dar caladas a un cigarro o gesticular con tics nerviosos. Todas estas acciones poseen características comunes: los episodios sucesivos son relativamente breves y el tiempo invertido en su ejecución es más o menos el mismo en todas las ocasiones.

Duración: la frecuencia o tasa de respuesta es una medida habitual de la magnitud del comportamiento, y otra es la duración relativa de una conducta, o más específicamente, la suma de sus duraciones divididas entre el tiempo total, también importante para estimar su cantidad. La *duración relativa* de la conducta es equivalente a la duración del tiempo que ocurre a lo largo de un periodo concreto. Al abordar un comportamiento como las pataletas, podría ser más interesante estimar la duración que su frecuencia. De hecho, la frecuencia puede ser bastante ambigua al aplicarla a algo como las pataletas (Pear, 2004). ¿Qué debería constituir y computarse como respuesta individual? ¿Cada grito, llanto o patada al suelo? ¿O quizá debiéramos compu-

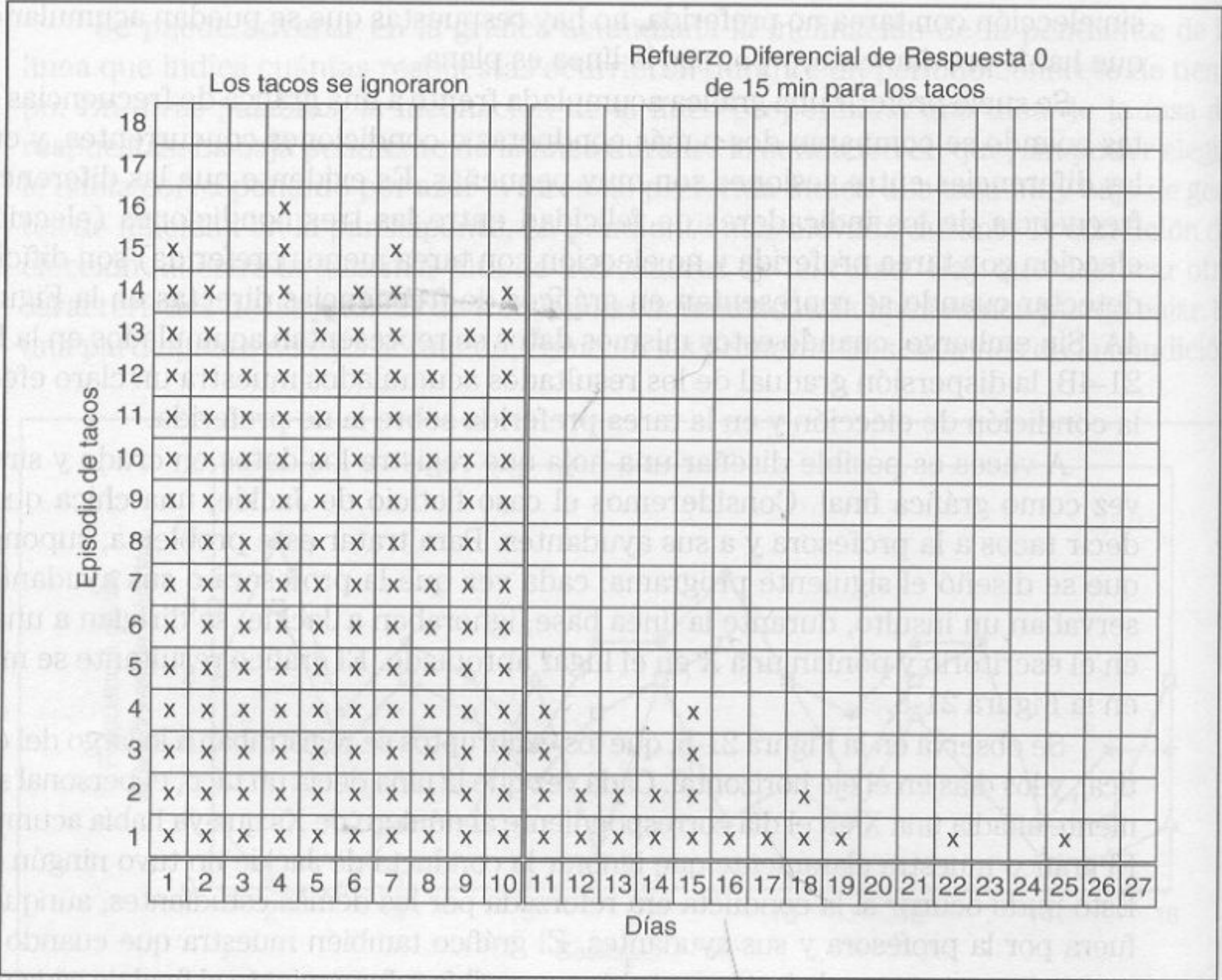


Figura 21-5. Conducta de decir tacos de Jackie. Cada X representa un taco. (Nota: estos datos son hipotéticos y se presentan sólo para ilustrar el procedimiento de representación). Este tipo de gráfico es especialmente útil para aquellos que no disponen de tiempo para volver a representar las frecuencias y pasar los datos de la tabla original a un gráfico más elaborado.

tar todos los episodios de berrinches como respuestas separadas? Es difícil responder a estas preguntas, sin embargo, por lo general las esquivamos si nos centramos sobre la duración de las pataletas. Otros ejemplos de conductas para las que registrar la duración de la respuesta es más apropiado que calcular la frecuencia serían escuchar con atención, permanecer sentada en su sitio en clase, ver televisión, hablar por teléfono y tomarse descansitos para café.

Si sólo le interesa tener un registro de la duración relativa de alguna actividad en varios días o sesiones sucesivas, es relativamente sencillo lograr una presentación visual muy clara mediante su tabulación y presentación en un gráfico combinado con la hoja de datos. Por ejemplo, una persona que quiere hacer un seguimiento de su conducta de ver televisión puede preparar un cuadro mostrando los minutos acumulados de televisión en el eje vertical y los días en el eje horizontal. La pendiente de esta gráfica indicaría la duración relativa del tiempo que pasa frente al televisor, de la misma forma que la pendiente del gráfico de frecuencia acumulada indicaba la tasa de respuesta. Se suelen usar cronómetros o relojes para registrar el tiempo.

Intensidad

A veces lo que interesa es medir la fuerza o intensidad de una respuesta y para ello suelen emplearse distintos instrumentos. Por ejemplo, para estimar el volumen de voz, medimos el nivel de decibelios mediante un *decibelímetro*. Para medir la fuerza con que se hace presión, por ejemplo en un apretón de manos, se puede emplear un *dinamómetro*. Las medidas de fuerza son comunes en las habilidades precisas para la práctica de varios deportes. En la actualidad existen aparatos que miden la fuerza con que lanza la pelota los pitcher de béisbol, o disparan el disco los jugadores de jockey, a partir de la velocidad que alcanzan estos objetos.

Control de los estímulos

A menudo nos gustaría evaluar una conducta en términos de las condiciones bajo las que se observaría su aparición. Como se señaló en el Capítulo 8, el término *control de los estímulos* se usa para indicar que una actividad específica ocurre en presencia de determinados estímulos y no aparece en presencia de otros. Hardy y sus colaboradores (1981) diseñaron el sistema de Evaluación Conductual Objetiva de los Discapacitados Mentales Severos y Moderados (*Objective Behavioral Assessment of the Severely and Moderately Mentally Handicapped, OBA*), para evaluar el control que ejercen los estímulos sobre las habilidades básicas de cuidado personal, habilidades sociales y de manejo personal más complejas, estrategias domésticas, destrezas motoras, y rendimiento en trabajos de formación profesional tutelados, de personas con trastornos de desarrollo severos y moderados. En esta prueba, los participantes reciben instrucciones para realizar una conducta concreta, por ejemplo «por favor, ponte los calcetines». La ejecución de la actividad se puntúa posteriormente conforme muestra la Tabla 21-2.

Tabla 21-2. Puntuación en la prueba de evaluación conductual objetiva de los discapacitados mentales severos y moderados (oba)	
Actividad de la prueba	puntuación
La actividad se realizó apropiadamente en todos los aspectos sin instigación ni guías de ningún tipo después de la presentación de las instrucciones específicas	3
La actividad se realizó correctamente sólo después de las instrucciones y de una instigación verbal por parte del evaluador	2
La actividad se realizó correctamente tras las instrucciones y una instigación verbal prescriptiva (similar a la mera instigación verbal salvo por el hecho de que incluye mucho más detalle).	1
La actividad no se realizó correctamente tras el nivel de instigación previo	0

Hardy y su equipo (1981) identificaron inicialmente conductas específicas que se practican en muchos programas de entrenamiento con personas con trastornos del desarrollo severos y moderados. Posteriormente, especificaron estos comportamientos en la prueba OBA; prepararon instrucciones para los evaluadores y estandarizaron definiciones de los diferentes tipos de instigadores, de forma que las actividades pu-

dieran evaluarse sobre la base del sistema de puntuación presentado en la Tabla 21-2. Este sistema de evaluación para identificar las condiciones bajo las que ocurrirá una conducta resulta muy útil para evaluar y poder ubicar a los estudiantes en programas de entrenamiento individualizados.

En muchos casos, los programas de modificación de conducta en que se trabaja para desarrollar habilidades pre-verbales y verbales van precedidos por evaluaciones conductuales del control que ejercen los estímulos sobre el rendimiento lingüístico de los estudiantes. Existen pruebas que determinan las condiciones bajo las que los alumnos emitirán conductas de petición, ecoicas o de denominación de objetos (conforme a lo descrito en el Capítulo 19, véase también Marion y cols., 2003). A este respecto, podríamos considerar que cualquier prueba en que los participantes reciben papel y lápiz, e instrucciones para responder a las preguntas, constituirá una prueba de control de los estímulos (véase **Nota 2**). ¿Están las respuestas correctas bajo el control de las preguntas? En muchos programas de entrenamiento la medida crítica de la conducta consiste en valorar si los estudiantes identifican correctamente los estímulos en forma de imágenes o palabras. En estos casos, la respuesta de denominación adecuada de los estudiantes está controlada por los estímulos que están identificando.

Latencia

Otra característica de la conducta que recibe atención es la latencia: el tiempo transcurrido entre la ocurrencia de un estímulo y el principio de una respuesta. Por ejemplo, una alumna podría trabajar de manera efectiva en clase una vez que empieza; el problema radica en la latencia de respuesta; es decir, cuando la profesora le pide que haga algo, «da vueltas y más vueltas» antes de empezar. Como sucedía con la duración, la latencia también se estima mediante cronómetros o relojes.

Calidad

La preocupación por la calidad de la conducta se observa con frecuencia en la vida cotidiana. Los profesores describen la calidad de la caligrafía de los escolares como «buena», «en la media», o «pobre». En deportes con jurado como el salto de trampolín, la gimnasia o el patinaje artístico, los atletas reciben puntos de acuerdo con la calidad de sus actuaciones. En nuestras vidas, es habitual que nos marquemos como objetivo hacer algunas actividades «mejor». Pero la calidad no es una característica adicional a las mencionadas previamente; más bien, es un refinamiento de una o más de las anteriores cualidades del comportamiento. A veces las diferencias en los juicios de calidad se basan en la topografía, como cuando el salto de una patinadora artística que aterriza apoyando sobre un pie se considera mejor que uno en que aterriza sobre dos pies. Respecto a la frecuencia y al control de los estímulos, muchas evaluaciones generales acerca de si una persona es buena o mala en una tarea, se relacionan con las veces que tiende a emitir una acción apropiadamente en un periodo de tiempo. Por ejemplo, una persona que es buena en los estudios, es seguramente alguien que muestra una alta frecuencia de estudio y de respuestas correctas en los exámenes. Los niños de quien se dice que son «buenos», muestran una frecuencia alta de «obe-

diencia»; esto es, obedecen las instrucciones de madres, padres y profesores. En términos de latencia, el caso de la corredora que sale de los tacos muy rápidamente después de oír el pistoletazo de salida constituye un «buen» comienzo, mientras que los corredores con una latencia más larga habrán ejecutado una salida «pobre». La calidad de la respuesta es esencialmente una designación arbitraria de una o varias de las características de la conducta mencionadas previamente, que se identifican como portadoras de algún valor funcional.

ESTRATEGIAS PARA REGISTRAR LA CONDUCTA

Podemos intentar registrar una conducta final ya definida en cualquier momento en que la persona tenga una oportunidad de emitirla. En la mayoría de los casos este método es demasiado ambicioso para el tiempo y recursos de que disponemos. Una alternativa es diseñar un segmento de tiempo específico, como una sesión de entrenamiento de una hora, una tarde, una comida o un tiempo de receso, e intentar registrar los episodios de la conducta pertinente durante ese intervalo. El *registro continuo* recoge todas las apariciones del comportamiento durante un periodo de tiempo específico.

Una estrategia alternativa es el *registro por intervalos*. Aquí, se selecciona un bloque específico de tiempo; por ejemplo, un periodo de observación de 30 minutos. Este periodo se divide en intervalos iguales de duración relativamente corta (p.ej., intervalos de diez segundos) y se registra la aparición de la conducta durante los mismos. Hay dos tipos de procedimiento de registro de intervalos. El más habitual se denomina *registro de intervalo parcial*, en que se registra la aparición de la conducta de manera dicotómica (se ha producido o no se ha producido), sin tener en cuenta la cantidad de veces que haya ocurrido durante el intervalo, ni su duración. Una observadora podría emplear una grabadora que reproduzca un pitido grabado previamente (o alguna señal) cada diez segundos. Supongamos que la conducta que interesa observar es una respuesta de interacción social definida. Si la respuesta ocurre una vez en un intervalo de diez segundos, se hace una marca en la hoja de registro (para ver un ejemplo de una hoja de recogida de datos, véase Figura 21-6). Si ocurren varias respuestas o una interacción social continuada durante el intervalo de diez segundos, la observadora hace no obstante sólo una marca. Tan pronto como suene el pitido, indicando el inicio del siguiente intervalo de diez segundos, la conducta se vuelve a marcar como 1 ó 0, dependiendo de su ocurrencia o no. Alternativamente, se puede codificar conforme al segundo tipo de procedimiento, el *registro de intervalo completo*. Con este acercamiento, se registra la conducta de interés sólo si persiste durante el intervalo de tiempo completo. La conducta registrada con cualquiera de los dos métodos se representa, típicamente, en términos de porcentaje de intervalos en que ha ocurrido.

Otra técnica de registro es el *registro de muestreo temporal* (véase, p.ej., Powell, Martindale y Kulp, 1975), en que se puntúa una conducta como presente o ausente durante intervalos de tiempo muy breves que están separados entre sí por periodos de tiempo mucho más largos. Por ejemplo, al padre o la madre de un

<u>Observador 1</u>		Intervalos de observación (10 segundos cada uno)																			
		5	10	15	20	25	30	35	40	45											
Tocar		/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	18	
Vocalizar		/	/	/				/	/	/	/	/	/	/	/	/	/	/	/		

<u>Observador 2</u>		Intervalos de observación (10 segundos cada uno)																			
		5	10	15	20	25	30	35	40	45											
Tocar		/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	18	
Vocalizar		/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/		

Figura 21-6. Ejemplo de hoja de datos para el registro por intervalos.

niño de preescolar le podría interesar con qué frecuencia se sienta y se mece su hijo (una conducta auto-estimulante). Quizá fuera útil tener registros de estas acciones, cuándo y cuánto suceden a lo largo de las horas de vigilia del niño, pero en general no es realista pensar en semejante registro. Una alternativa sería que el padre buscara y observara al hijo una vez por hora y que anotara si el niño se mece durante un intervalo de observación de quince segundos. Cada intervalo de observación estaría separado del siguiente por una hora aproximadamente. Este tipo de técnica de observación permite registrar una o más conductas de uno o más alumnos, aunque los observadores tengan otras muchas ocupaciones a lo largo del día. Un ejemplo de hoja de registro de muestreo temporal aparece en la Figura de 21-7. Una variación de éste se denomina *muestreo temporal momentáneo* y consiste en el registro dicotómico (ocurre o no ocurre) de la actividad en momentos concretos como, por ejemplo, a las horas en punto.

A menudo se usan procedimientos de registro con características de los registros por intervalos y de muestreo temporal simultáneamente. Se podría vigilar al aprendiz durante un intervalo específico (digamos diez segundos) y anotar las actividades durante los diez segundos siguientes. Esta estrategia de *observación* (durante diez segundos) y *registro* (durante diez segundos) continuaría durante un periodo de tiempo (por ejemplo, media hora). De esta forma, una sola observadora podría mirar a un estudiante durante diez segundos y registrar después si la conducta ha ocurrido o no, atender luego a otro estudiante durante diez segundos y registrar después si la conducta se ha producido o no, y seguir así, hasta que se hubiera observado y registrado en una ocasión el comportamiento de todos los alumnos. Luego procedería a realizar una segunda ronda y una tercera y así hasta completar el periodo total de observación. Este procedimiento se conoce como registro por intervalos, aunque, estrictamente hablando, también podría ser descrito como muestreo temporal con periodos muy breves entre los intervalos de observación.

Fecha _____	Conducta			Lugar			Comentarios
	Sentado	De pie	Balanceándose	Cocina	Salón	Dormitorio	
Hora							
8:00 AM							
9:00							
10:00							
11:00							
12:00 PM							
1:00							
2:00							
3:00							
4:00							
5:00							
6:00							
7:00							
8:00							
9:00 PM							

Figura 21-7. Ejemplo de hoja de datos de muestreo temporal para un niño que se balancea frecuentemente mientras está sentado.

EVALUACIÓN DE LA PRECISIÓN DE LAS OBSERVACIONES

Hawkins y Dotson (1975) identificaron tres fuentes de error que afectan a la precisión de las observaciones. En primer lugar, la *definición de la respuesta* puede ser vaga, subjetiva o incompleta, de forma que los observadores tendrían problemas para hacer registros precisos. En segundo lugar, la *situación observacional* podría dificultar la detección de la actividad debido a distractores u otras obstrucciones al proceso, o debido a que la conducta sea demasiado sutil o compleja para ser analizada con precisión en esa situación. En tercer lugar, los *observadores* pueden estar poco entrenados, poco motivados, mal predispuestos o ser incompetentes. Se podrían añadir dos posibles fuentes de error: *hojas de registro* mal diseñadas o un *procedimiento de registro* demasiado engorroso. Debido a que una o la combinación de varias fuentes de error pueden estar presentes en cualquier proyecto de modificación de conducta, los investigadores suelen llevar a cabo estimaciones de **fiabilidad entre observadores o jueces (FIO)**. Dos observadores independientes podrían registrar observaciones de la misma conducta de la misma persona durante una sesión concreta, y ser

cuidadosos para no influir ni enviarse señales entre sí mientras registran, o copiarse mutuamente las observaciones. La pregunta que podríamos hacernos es, teniendo en cuenta los esfuerzos necesarios para observar mientras usan las definiciones de conducta disponibles y los procedimientos de registro, y considerando su entrenamiento, ¿cuánto se parecerán sus observaciones? Hay varias formas de averiguarlo, pero existen dos procedimientos para calcular la fiabilidad entre jueces frecuentemente utilizados.

Un procedimiento para calcular, de fiabilidad entre observadores, se ilustra de la siguiente manera. Volvamos al ejemplo del observador que está registrando el número de elementos (saltos y giros) de una patinadora artística, tal como los definió la entrenadora previamente en este capítulo. Imaginemos que traemos un segundo observador, que se sitúa en el lado opuesto de la pista y observa también a la patinadora. El segundo observador está familiarizado con la definición de salto y giro y emplea una hoja de registro idéntica a la del primer observador. Al final de la sesión, el primer observador registró veinte elementos y el segundo registró veintidós elementos. Para convertir estos datos en una estimación de la fiabilidad entre jueces, dividimos el número menor entre el número mayor y multiplicamos por 100%: la fiabilidad obtenida es de 91%. Este porcentaje de fiabilidad significa que los dos jueces estaban muy de acuerdo (casi 100%) sobre el *número* total de elementos. No significa que hayan estado de acuerdo en veinte elementos específicos y que el segundo observador haya contado dos elementos más. Es muy posible que un observador haya registrado un elemento que al segundo se le pasara por alto, o éste contar un elemento que el anterior no viera. Si estas diferencias se hubieran producido a lo largo de todo el entrenamiento, los observadores no habrían coincidido en ninguna respuesta individual específica. No obstante, un porcentaje de acuerdo tan elevado proporciona la confianza de que el número total de elementos ocurridos se acerca mucho al número que cada observador registró, a pesar de algunos posibles desacuerdos. Esta aproximación de contar dos totales y luego dividirlos entre sí y multiplicar por 100 es muy habitual cuando dos observadores cuentan la frecuencia de una respuesta particular durante un periodo de tiempo concreto.

Otro procedimiento para calcular la fiabilidad entre jueces es el registro por intervalos. Recordemos que en los procedimientos de registro por intervalos, una y sólo una respuesta se registra durante un periodo de tiempo breve (normalmente cinco o diez segundos) sobre un periodo extendido de observación. Si tenemos dos observadores independientes registrando la misma conducta, y cada uno está empleando un procedimiento de registro por intervalos, ¿cómo se comparan sus intervalos sucesivos en relación con aquellos que contienen la respuesta frente a los que no la contienen? Supongamos que los dos observadores están registrando dos tipos de interacción social en un chico. Las conductas están definidas como tocar a otro chico y vocalizar en la dirección de otro niño. Sus puntuaciones de intervalos se muestran en la Figura 21-6.

Queda claro que ambos observadores anotaron dieciocho episodios de contacto físico; no obstante, los dos observadores coincidieron sólo en dieciséis de los dieciocho episodios. Cada uno marcó dos episodios que pasaron desapercibidos para la otra persona, lo cual arroja un total de cuatro desacuerdos. Si hubiéramos aplicado el procedimiento descrito previamente, habríamos obtenido una fiabilidad entre jueces del

100%. Sin embargo, en el segundo procedimiento, el índice de fiabilidad entre los observadores se obtiene dividiendo el número de intervalos en que los dos observadores coincidieron en que la conducta aparecía, entre el número total de intervalos en que sólo uno de los dos la registró (los acuerdos divididos entre acuerdos y desacuerdos) y multiplicando por 100%. De esta forma, el segundo procedimiento daría una fiabilidad entre observadores de 89%.

Típicamente, por convención, en los estudios de modificación de conducta, se consideran aceptables las puntuaciones de fiabilidad entre observadores que se sitúan entre 80% y 100%. La variación potencial en los procedimientos de registro y cálculo, no obstante, podría invalidar la estimación de fiabilidad obtenida o llevar a conclusiones erróneas si la consideráramos aisladamente. Sugerimos por ello a los lectores que consideren las definiciones de la respuesta, los procedimientos de entrenamiento de los observadores, el sistema de registro, el método para calcular la fiabilidad entre jueces y el valor final de la misma, antes de evaluar la fiabilidad de los datos reseñados. Los defectos en alguno de estos aspectos debieran suscitar dudas acerca de los resultados.

PREGUNTAS DE REPASO

1. ¿Qué quiere decir el término topografía de la respuesta? Describa un ejemplo.
2. ¿Cuáles son las dos medidas de cantidad de la conducta? Defina y aporte un ejemplo de cada una.
3. Describa tres tipos de seguimiento del número de veces que una respuesta concreta ocurre a lo largo de un día.
4. Prepare un gráfico de frecuencias acumuladas de los siguientes episodios de una conducta que se observaron durante sesiones sucesivas: 3, 7, 19, 0, 0, 0, 27, 12, 12, 6.
5. Describa al menos cuatro formas en que una gráfica acumulada de un conjunto de datos difiera de una gráfica de frecuencias con los mismos datos.
6. En una gráfica de frecuencias acumuladas, ¿qué se puede inferir de lo siguiente?
 - a. una pendiente pronunciada (muy inclinada)
 - b. una pendiente baja (poco inclinada)
 - c. una línea plana
7. ¿Qué dos características suelen mostrar las conductas registradas en términos de frecuencias?
8. ¿Qué significa duración relativa de una conducta? Aporte y explique un ejemplo en que la duración relativa sea una estimación más apropiada que la frecuencia.
9. ¿Qué otro término se aplica para denominar la intensidad de la respuesta? Describa un ejemplo en que sea importante medir la intensidad de una conducta.
10. ¿Qué significa el término latencia de la respuesta? Aporte un ejemplo.
11. ¿Qué evalúa el sistema de Evaluación Conductual Objetiva de los Discapacitados Mentales Severos y Moderados (*Objective Behavioral Assessment of the Severely and Moderately Mentally Handicapped, OBA*)? Explique su respuesta.
12. ¿Qué características conductuales se usan para evaluar la calidad de una respuesta? Justifique mediante ejemplos.
13. Describa con un ejemplo el sistema de registro continuo.
14. Describa con un ejemplo el sistema de registro de intervalo parcial.
15. Describa con un ejemplo el sistema de registro de muestreo temporal.

16. ¿Cuándo es probable que se prefiera un sistema de registro por intervalos frente a un sistema de registro continuo?
17. Describa cinco fuentes de error en el registro de observaciones.
18. Explique en una frase o dos qué quiere decir fiabilidad entre observadores. Describalo mediante palabras, no mediante los procedimientos para calcularla.
19. Mediante el procedimiento descrito en el texto para calcular la fiabilidad entre observadores con datos de intervalos, calcule las estimaciones pertinentes para los datos de vocalizaciones registradas por los observadores 1 y 2 (Figura 21-6). Muestre sus cálculos.
20. De acuerdo con la convención, ¿cuál es la fiabilidad entre observadores aceptable en un programa de investigación? ¿Qué quiere decir «por convención»?

EJERCICIOS DE APLICACIÓN

A. Ejercicio que implica a otras personas

Seleccione un déficit o exceso conductual que se haya modificado satisfactoriamente (p.ej., las pataletas de Peter), tal como se describe en alguno de los otros capítulos. Para esa conducta:

1. Diseñe una hoja de registro plausible, incluyendo una columna para sesiones y otra columna para los episodios de la conducta por sesión.
2. Prepare un resumen de algunos datos

representativos (reales o hipotéticos) y anótelos en la hoja de registro.

3. Haga una gráfica de frecuencias con los datos.
4. Haga una gráfica de frecuencias acumuladas con esos datos.

B. Ejercicio de auto-modificación

1. Seleccione uno de sus propios excesos o déficit conductuales. Responda las preguntas 1 a 4 del ejercicio anterior para esa conducta.

NOTAS DE AMPLIACIÓN SOBRE EL TEMA

1. Cuando el método de observación afecta a las conductas que se analizan, se dice que la observación es *intrusiva*. Registrar observaciones de *manera no intrusiva* significa que no deberían causar alteraciones en el comportamiento típico de quien se observa. En otras palabras, no se persigue que las observaciones afecten a la conducta que se observa, y por ello existen técnicas para asegurarse de que las actuaciones no tendrán carácter intrusivo. Una posibilidad es observar la conducta a través de un cristal unidireccional, como se ilustra en el caso de Darren del Capítulo 3. Otra posibilidad es observar a los individuos a distancia, sin que se den cuenta. Esta estrategia se usó para estudiar los há-

bitos de bebida de los clientes habituales de los bares de una ciudad media estadounidense (Sommer, 1977). Otro método es tener un cómplice que haga las observaciones mientras trabaja mano a mano con los clientes en su ambiente normal de trabajo (Rae, Martin y Smyk, 1990). Otras alternativas incluyen grabaciones con cámara oculta y la evaluación de los productos de las actividades de los usuarios (por ejemplo, codificar los objetos arrojados a la basura en una zona pública de acampada; Osborne y Powers, 1980). Sin embargo, tales tácticas plantean un dilema moral: ¿es ético observar a las personas sin su consentimiento? La Asociación Estadounidense de Psicología (*American Psychological*

Association (APA), 2000) ha desarrollado una serie de pautas éticas respecto a todos los tipos de experimentos de investigación psicológica. Entre los factores que estipulan estas pautas para hacer observaciones con propósitos experimentales se encuentran: contar con el consentimiento de la persona observada o de su tutor legal; que las observaciones contribuyan al programa de tratamiento de la persona observada; que se garantice el mantenimiento de la confidencialidad de las observaciones; que las personas concretas o grupos no sean reconocibles en los informes que surjan de las observaciones. Cualquiera que se plantee registrar la conducta de otra persona debiera consultar las pautas éticas de su organización profesional y las leyes aplicables sobre privacidad y confidencialidad. (Véase también el Capítulo 30 de este libro).

2. Otro instrumento de evaluación del control de los estímulos es la prueba ABLA, Evaluación de las Habilidades Básicas de Aprendizaje (*Assessment of Basic Learning Abilities*), desarrollada por los pioneros conductistas Nancy Kerr y Lee Meyerson. EL ABLA evalúa la facilidad o dificultad con que las personas con trastornos del desarrollo profundos, severos o moderados son capaces de aprender tareas en un rango de dificultad que abarca desde la simple imitación, hasta discriminaciones relativamente complejas con claves visuales o auditivas. El nivel 3, por ejemplo, es una tarea de discriminación visual con dos opciones en que se pide a los participantes que coloquen una pieza de espuma en una lata amarilla que alterna aleatoriamente su posición con una caja roja. El nivel 4, es una discriminación de identidad visual-visual en que los participantes reciben aleatoriamente un cubo rojo o un cilindro amarillo para colocarlo en la caja roja o en la lata amarilla, que alternan posiciones (izquierda y derecha). El nivel 6 es una discriminación auditiva-visual de dos opciones, en que los evaluadores dicen al azar «Caja roja» o «Lata amarilla» y la respuesta correcta implica colocar la

pieza de espuma en el contenedor identificado cuando los dos contenedores acaban de alternar su posición a la vista de los participantes. Durante la evaluación de cada nivel del ABLA, las respuestas correctas se refuerzan y las respuestas incorrectas van seguidas por un procedimiento de corrección del error. Las pruebas en un nivel continúan hasta que los clientes superan el criterio de ocho respuestas correctas seguidas, o un fallo de ocho errores seguidos. La investigación sobre las pruebas del ABLA indica que los seis niveles están ordenados jerárquicamente conforme a la dificultad creciente de las mismas, y que el ABLA es una herramienta valiosa para profesores y rehabilitadores, porque les permite elegir y secuenciar el entrenamiento y las tareas para personas con trastornos del desarrollo profundos, severos y moderados (Martin y Yu, 2000; Martin, Yu y Vause, 2004). Por ejemplo, las personas que superan el nivel 4 del ABLA, son capaces de aprender rápidamente tareas de emparejamiento útiles socialmente, como poner los calcetines por parejas o reponer la barra de ensaladas de un restaurante de comida rápida.

3. El procedimiento que hemos sugerido para calcular la fiabilidad entre jueces durante el registro por intervalos es dividir el número de intervalos en que ambos observadores están de acuerdo en que una conducta ocurrió, entre el número total de intervalos en que alguno de ellos registró la conducta (acuerdos más desacuerdos) y multiplicarlo por 100%. Algunos investigadores, sin embargo, incluyen en sus medidas de acuerdos aquellas coincidencias de no ocurrencia de la conducta; en otras palabras, acuerdos en intervalos vacíos. Cuando se han registrado muy pocas conductas, no obstante, esto puede inflar bastante la puntuación de fiabilidad. Consideremos por ejemplo los 45 intervalos de observación que se muestran en la Figura 21-6. Supongamos que el observador 1 hubiera registrado un episodio de tocar durante el intervalo 5, que el observador 2 hubiera registrado un episodio durante

el intervalo 6 y que no se registraran más episodios de esta conducta. En este caso, los dos observadores estarían en total desacuerdo acerca de la ocurrencia de la conducta y la fiabilidad entre jueces sería cero si se calculara como sugerimos en el texto. Sin embargo, si los acuerdos en los intervalos vacíos se incluyeran, la fiabilidad entre jueces sería igual a 43 acuerdos divididos entre 43 acuerdos más dos desacuerdos multiplicado por 100%, lo que arrojaría un resultado de 95,6%. Debido a esta defor-

mación, muchos investigadores no cuentan los acuerdos en los intervalos vacíos. En otras palabras, los intervalos en que ningún observador registra la conducta se ignoran. Una excepción aceptable se produce cuando lo interesante es reducir una conducta y es importante coincidir acerca de si la conducta no tuvo lugar. Estos puntos y otros comentarios sobre la complejidad de calcular la fiabilidad entre observadores se discuten con más detalle en Poling, Methot y LeSage (1995).

Preguntas sobre las notas

1. ¿Cuál es la diferencia entre las observaciones intrusivas y las no intrusivas?
2. ¿Cuándo es especialmente engañoso incluir los acuerdos en intervalos vacíos al calcular la fiabilidad entre observadores? Aporte un ejemplo.
3. ¿Cuándo podría ser aceptable incluir los acuerdos en intervalos vacíos en los cálculos de la fiabilidad entre jueces? ¿Por qué sería adecuado?
4. ¿Qué información proporciona la prueba ABLA?